

Identification and Characterization of a Polymorphic Receptor Kinase Gene Linked to the Self-Incompatibility Locus of *Arabidopsis lyrata*

Mikkel H. Schierup,* Barbara K. Mable,^{†,1} Philip Awadalla[†] and Deborah Charlesworth[†]

*Department of Ecology and Genetics, University of Aarhus, Ny Munkegade, DK-8000 Aarhus C, Denmark and [†]Institute of Cell, Animal and Population Biology, University of Edinburgh, Ashworth Laboratories, Edinburgh EH9 3JT, United Kingdom

Manuscript received October 16, 2000
Accepted for publication January 19, 2001

ABSTRACT

We study the segregation of variants of a putative self-incompatibility gene in *Arabidopsis lyrata*. This gene encodes a sequence that is homologous to the protein encoded by the *SRK* gene involved in self-incompatibility in Brassica species. We show by diallel pollinations of plants in several full-sib families that seven different sequences of the gene in *A. lyrata* are linked to different *S*-alleles, and segregation analysis in further sibships shows that four other sequences behave as allelic to these. The family data on incompatibility provide evidence for dominance classes among the *S*-alleles, as expected for a sporophytic SI system. We observe no division into pollen-dominant and pollen-recessive classes of alleles as has been found in Brassica, but our alleles fall into at least three dominance classes in both pollen and stigma expression. The diversity among sequences of the *A. lyrata* putative *S*-alleles is greater than among the published Brassica *SRK* sequences, and, unlike Brassica, the alleles do not cluster into groups with similar dominance.

THE hope of elucidating the complete molecular determination of a self-incompatibility (SI) system in plants has grown significantly with recent investigations of the sporophytic incompatibility system in the genus Brassica. Male (pollen) and female (stigma) components of the recognition/incompatibility reaction appear to be controlled by separate genes that reside in a small genomic region (the *S*-locus; see YU *et al.* 1996; SCHOPFER *et al.* 1999). The interaction between male and female components is not completely understood, but it is thought that a pollen surface protein acts as a ligand that is recognized by a transmembrane protein in the papillary cells on the surface of the stigma. When the pollen and pistil specificities are from the same *S*-allele, pollen tube growth is inhibited. The stigma component of this recognition system is now thought to be the *S*-locus receptor kinase, encoded by the *SRK* gene. This protein has an extracellular glycoprotein domain and an intracellular serine-threonine protein kinase (STEIN *et al.* 1991) and has been shown to be necessary, and perhaps sufficient, for determining specificity (CUI *et al.* 2000; TAKASAKI *et al.* 2000). A second protein, *S*-locus glycoprotein, encoded by the closely linked *SLG* gene, is not in itself sufficient for determining specificity, although it may be necessary for proper rejection of incompatible pollen (SHIBA *et al.* 2000; TAKASAKI *et al.* 2000). *SLG* sequences show homology

to those of the first exon of *SRK* (the *S*-domain). A pollen coat protein, encoded by the linked *SCR* gene, has recently been shown to be necessary and sufficient for determination of the pollen specificity (SCHOPFER *et al.* 1999; TAKAYAMA *et al.* 2000).

In Brassica, the *SRK*, *SLG*, and *SCR* genes are located close to one another in the physical map. All display very high levels of sequence diversity (NISHIO and KUSABA 2000; WATANABE *et al.* 2000). At least 36 *SLG* alleles have been characterized so far, with pairwise differences among alleles ranging from 2–40% at the amino acid level (KUSABA *et al.* 1997; NISHIO and KUSABA 2000). On the basis of the *S*-domain sequences, *SLG* and *SRK* alleles from *Brassica oleracea* and *B. rapa/campestrifera* are intermingled in gene trees, suggesting that most of the polymorphism predates the divergence of these two species (NASRALLAH *et al.* 1987; DWYER *et al.* 1991; KUSABA *et al.* 1997). These observations are in agreement with population genetics theory, which predicts that self-incompatibility alleles should be maintained in the population for very long times due to the action of frequency-dependent selection (TAKAHATA 1990; VEKEMANS and SLATKIN 1994; SCHIERUP *et al.* 1998). The *S*-domain sequences display three hypervariable (HV) regions, which are suspected to be the regions involved in recognition, and thus the main targets of the selection that maintains the variability (SIMS 1993; AWADALLA and CHARLESWORTH 1999; NISHIO and KUSABA 2000). This remains controversial, however, as the HV regions could merely be regions of relaxed selective constraint (CHARLESWORTH *et al.* 2000). The *S*-domain sequences of the two pistil-expressed genes (*SLG* and *SRK*) from a given *S*-allele are on average

Corresponding author: Mikkel H. Schierup, Department of Ecology and Genetics, University of Aarhus, Ny Munkegade, Bygn. 540, DK-8000 Aarhus C, Denmark. E-mail: mikkel.schierup@biology.au.dk

¹Present address: Department of Botany, University of Guelph, Guelph, Ontario N1G 2W1, Canada.

more closely related to each other than those between alleles of different specificities, so that the variability displays haplotype structure (KUSABA *et al.* 1997; AWADALLA and CHARLESWORTH 1999; NISHIO and KUSABA 2000). Again, this is expected for neutral variants that are linked to a site or sites in a gene that is subject to balancing selection (*e.g.*, STROBECK 1983; VEKEMANS and SLATKIN 1994; NORDBORG *et al.* 1996; TAKAHATA and SATTA 1998; SCHIERUP *et al.* 2000) and has been documented in mammalian MHC sequence data (*e.g.*, HUGHES and YEAGER 1998).

In sporophytic self-incompatibility systems, dominance is possible between pairs of alleles in the determination of the pollen phenotype. In Brassica, pollen-dominant and -recessive *S*-alleles are called class I and II, respectively. The *SLG*/*SRK* sequences of dominant and recessive alleles are very distinct (NASRALLAH *et al.* 1991). Pairwise amino acid sequence differences between *S*-domains of class I and II *SLGs* and *SRKs* average ~35% (NASRALLAH *et al.* 1991), compared with ~20% differences within either class (KUSABA *et al.* 1997), and class I and II alleles differ so greatly that alleles of one class are often not recognized by probes for the other (KUSABA *et al.* 1997; NISHIO and KUSABA 2000). Of the *S* haplotypes so far identified, most have been classified as class I, whereas only three fall into class II, so estimates of sequence differences are subject to considerable error.

Although understanding of the molecular control of SI in Brassica has progressed significantly in recent years, there are many unanswered questions that could benefit from characterization of additional sporophytic systems. Some of the most intriguing questions are how such high diversity at a multigene locus can be maintained and how coordinated changes in pollen and pistil specificity are possible (*e.g.*, CHARLESWORTH 2000). These questions require an evolutionary perspective and can be studied in natural populations, where the number of alleles, the diversity within and between alleles, and the molecular nature of polymorphisms can be characterized using population genetics approaches.

In addition to the work on Brassica outlined above, the orthologue of *SLG* has been investigated in another cultivated plant, *Raphanus sativus* (SAKAMOTO *et al.* 1998), which is closely related to Brassica. Both are members of the tribe Brassicaceae (ROLLINS 1993) and viable hybrids are known between members of the two genera (including the classic example *Raphanobrassica*; see KARPECHENKO 1927; McNAUGHTON 1973). The close relatedness is supported by comparisons of sequences of internal transcribed spacer (ITS) regions of rDNA and the mitochondrial NADH gene (YANG *et al.* 1999a,b). Eighteen *R. sativus* *S*-domain sequences (presumed to be *SLG*) all share the 12 cysteine residues that are conserved among Brassica *S*-domains, and the hypervariable sequences are found in the same regions as in Brassica. The *R. sativus* putative *SLG* alleles could

also be categorized into class I and II sequences, on the basis of their relative similarity to the corresponding Brassica alleles. Most polymorphism thus appears to predate the split of these closely related genera. *SRKs* have not yet been characterized from *R. sativus*.

We recently began a search for orthologues of *SRK* and *SLG* in *Arabidopsis lyrata* (also called *Cardaminopsis petraea*, *Arabis petraea*, or *Arabis lyrata*). *A. lyrata* is one of the closest relatives of *A. thaliana*, but is self-incompatible. It is a perennial herb with a circumboreal distribution. The genus *Arabidopsis* is a member of the tribe Arabidae (ROLLINS 1993) and it is much more distantly related to Brassica than is *Raphanus*, based on ITS and NADH sequences (YANG *et al.* 1999a,b).

Using a distantly related naturally occurring species not only allows evaluation of the nature of SI under noncultivated conditions (and thus allows study of the effects of population structure of the variants and comparison with reference loci that are not expected to be under balancing selection; *e.g.*, see SCHIERUP *et al.* 2000) but also allows investigation of whether the SI system is homologous within the family Brassicaceae. By understanding the SI system in a close relative of the self-compatible plant *A. thaliana*, it may also be possible to determine how its self-compatibility evolved. *A. thaliana* could have lost self-incompatibility either because the *S*-locus region had been deleted, or, alternatively, orthologues of the *S*-loci may remain in *A. thaliana*, but no longer confer self-incompatibility. Comparative mapping with Brassica suggests that the genomic region containing the pistil genes of the *S*-locus is deleted in *A. thaliana* (CONNER *et al.* 1998) but a closer taxonomic comparison should be done to test this further. Furthermore, transfer of the *S*-locus genes from self-incompatible to self-compatible species is being attempted to evaluate the set of genes from the self-incompatible species that is necessary and sufficient to produce an SI reaction in *A. thaliana*. Recently, the genomic region containing the *SRK*, *SLG*, and *SCR* genes, as well as a downstream gene, *ARCI*, was transferred from Brassica to the self-compatible *A. thaliana*, but this was not sufficient to produce a self-incompatibility reaction in *A. thaliana* (BI *et al.* 2000). Such experiments should be less problematic if the important regions are derived from *A. lyrata* rather than the more distantly related Brassica species, because the necessary downstream genes are more likely to be present.

We began this study with the following questions in mind:

1. Are both *SRK* and *SLG* required for the stigmatic incompatibility reaction in *A. lyrata*, or is one of them dispensable?
2. Are the same regions of the genes hypervariable in *A. lyrata* as in Brassica, and are these the regions that are important in determining specificity?

3. How old is the present polymorphism in the *SRK*/*SLG* genes within the Brassicaceae?
4. Is there a similar system of dominance in all species of Brassicaceae?

Our strategy was based on starting with sequence information from the *S*-domains of Brassica *SRK* and *SLG* alleles and designing primers to their most conserved regions. Because of the large evolutionary distance between the species, this strategy led, as expected, to amplification of several members of the *S* gene family in *A. lyrata* (CHARLESWORTH *et al.* 2000). Most of these genes either show very little polymorphism or are unlinked to the SI phenotype (CHARLESWORTH *et al.* 2000; D. CHARLESWORTH, P. AWADALLA, M. H. SCHIERUP and B. K. MABLE, unpublished results). However, one set of sequences ("Type 13" in CHARLESWORTH *et al.* 2000) showed very high levels of polymorphism, and preliminary evidence suggested that at least some of the putatively allelic variants were linked to SI. These sequences were thus identified as candidate alleles at the *S*-locus in *A. lyrata*.

Here we present evidence that sequences from this putative locus (here referred to as *Aly13*) may represent the *A. lyrata* *SLG* and *SRK* orthologues. We describe diallel pollinations in full-sib families and segregation analyses, showing that at least 11 of these sequences segregate as alleles at the same locus and that their sequences have a similar structure to the Brassica *SRK* gene. One of the alleles at this putative *S*-locus is almost identical in sequence to a stigma-expressed cDNA (*SRKa*) independently isolated by J. B. NASRALLAH and M. E. NASRALLAH and shown by them to have a kinase domain (personal communication). Furthermore, this locus is highly polymorphic, with more variability than the *SRK* in Brassica.

MATERIALS AND METHODS

Crossing and segregation analysis: Full-sib families were raised from crosses between known individuals originating from populations in Michigan, North Carolina, Scotland, and Iceland. Details of the populations studied are given in CHARLESWORTH *et al.* (2000). Seeds were planted in Fisons F2 compost with the addition of 30% 2–3 mm grit and raised in a greenhouse under a minimum of 12 hr light (with artificial light when necessary).

Six families were chosen for reciprocal diallel crosses. Plants to be pollinated were covered with net curtain fabric to exclude pollinators, and non-emasculated flowers were hand pollinated by rubbing dehisced anthers over their stigma. Each combination of parents tested was pollinated reciprocally with three replicate flowers. Compatibility was scored as fruit set about 7–10 days after pollination. Flowers where no fruits developed were classified as incompatible and flowers with full-sized fruits as compatible. In some cases, small fruits with one to two seeds developed; these were scored as small fruits because it is difficult to distinguish whether reduced fruit set is due to incompatibility at the pollen-stigma interaction level or whether it is due to incompatibility at a later stage in development that may be unrelated to *S*-locus specificity (*e.g.*,

MAHY and JACQUEMART 1999). The crossing results divided the plants into phenotypic classes. By definition, pollinations within a phenotypic class are incompatible and every individual within a phenotypic class shows the same pattern of incompatibility when tested with individuals of other classes. In a sporophytic SI system, families are expected to contain one to four phenotypic classes, depending on the number and dominance of self-incompatibility alleles of the parents. These phenotypic classes were subsequently compared with the genotypic classes defined by sequences at the putative self-incompatibility gene (hereafter termed *Aly13* "subtypes"; see below) to test for linkage. In all cases, crossing experiments were completed before genotypic typing, which thus could not influence the choice and interpretation of crosses.

In five more families, progeny were raised but not used for crosses. Instead, segregation of the *Aly13* subtypes present in the parents of the crosses was followed in the progeny to investigate whether different *Aly13* subtypes segregated as expected from alleles belonging to the same genetic locus.

DNA extraction and PCR condition: DNA was extracted from one to four fresh or frozen (–70°) leaves using a CTAB protocol (JUNGHANS and METZLAFF 1990). PCR conditions for the *S*-domain were the following: denaturation at 94° for 3 min followed by 35 cycles of 94° for 30 sec, 50°–54° for 1 min, and 72° for 1 min, followed by final extension at 72° for 5 min. Annealing temperatures were usually 50° for the general primers and 54° for the *Aly13* subtype-specific primers. For identification of kinase domains, gradient PCR was used in many cases with an annealing temperature gradient of 40°–55° and an extension time of 2 min. Normal PCR was performed using a PTC-200 thermal cycler (MJ Research, Watertown, MA) and gradient PCR using a PCRexpress thermal cycler (Hybaid).

Identification and sequencing of *Aly13* subtypes: The strategy for identifying *Aly13* sequence variants (subtypes) was outlined in CHARLESWORTH *et al.* (2000). Initial identification focused on the intron-free *S*-domain only. Candidate *SRK* and *SLG* orthologues were identified by restriction enzyme profiling (using *AhaI*, *RsaI*, and *MspI*; see CHARLESWORTH *et al.* 2000) of sequences amplified by the primers 13F1 (~300 bp from the beginning of the *S*-domain: 5' ccgacggtaacctgtcatcctc 3') and SLGR (~50 bp from the 3' end of the *S*-domain: 5' atcgacataagatcttgacc 3'); precise locations of the primers cannot be specified because different alleles are of different lengths. However, due to high divergence between different *Aly13* sequences, this primer set did not amplify all allelic sequences (*i.e.*, plants in certain incompatibility groups in our crosses were predicted to be heterozygous for two different *S*-alleles but only one, or sometimes no, *Aly13* sequence was found initially). We therefore used an alignment of the *S*-domain sequences from the *Aly13* sequence variants to design two further degenerate forward primers (13seq1F: 5' tgg aaa aa/gc tca/c tat gat cc 3' and 13seq2: 5' gat gga c/atc cgg/a tt ag/tc/t ggc at 3') located ~300 bp downstream of 13F1. Using a combination of these primers and 13F1 (with the reverse primer, SLGR) we were able to amplify sequences corresponding to most of the SI phenotypes determined from our crossing data (see below). All *Aly13* sequence subtypes were initially cloned from PCR products using TOPO TA cloning vectors (Invitrogen, San Diego). Subsequently, specific forward primers were designed for each sequence type, which aided identification of putative alleles. The subtype-specific primers were used for direct sequencing, whenever possible, to minimize possible errors due to misincorporation of bases in PCR prior to cloning. The sequence of these primers can be obtained from the authors by request.

The *Aly13* subtypes that showed evidence of linkage with SI phenotypes in our families were subsequently tested for

the presence of a kinase domain using the following strategy. The *S*-domain sequence of a stigma-expressed cDNA from *A. lyrata* (*SRKa*) obtained from J. B. NASRALLAH and M. E. NASRALLAH (personal communication) was compared with our *Aly13* sequences and found to differ by only a single nonsynonymous change from the *S*-domain of one of our sequence types (*Aly13-13*). The *Aly13-13* sequence was obtained from an individual collected from a population close to that from which the *SRKa* cDNA sequence was derived (Indiana Dunes). *SRKa* was then aligned with published Brassica *SRK* alleles and *A. thaliana* *S*-like receptor kinases, and several reverse primers from different exons of the kinase domain were designed based on regions conserved among these sequences (these primer sequences are available on request). These primers were used with either 13F1 or an *Aly13* subtype-specific primer to amplify products from genomic DNA of individuals carrying known *Aly13* subtypes. PCR bands resulting from these amplifications were subsequently cloned and sequenced. In some cases, initial amplification was weak due to the large size of products (2–3 kb), and reamplification from purified gel bands was necessary prior to cloning.

The resulting sequences were aligned with an alignment containing the full coding sequences of all 16 published Brassica *SRK* sequences, *SRKa*, and the *S*-domains from all previously identified *Aly13* types. This was critical both in the assessment of homology of the kinase domains and in confirming that the correct *Aly13* type had been amplified. It should be noted that while the procedure just described can confirm the presence of a kinase domain, negative results do not prove that the sequence types do not contain a kinase domain; it is possible that some sequences are too different to amplify using our primers. The large variation possible in the size of intron 1 (0.7–7 kb among Brassica *SRKs* (KUSABA *et al.* 1997) also increases the possibility that kinase domains may be missed using a PCR-based approach.

All sequencing was done on an ABI 377 automatic sequencing machine using either ThermoSequenaseII (Pharmacia, Piscataway, NJ) or BigDye (Applied Biosystems, Foster City, CA) sequencing technology. Sequences were subsequently checked manually for accurate base calling using SeqEd (Applied Biosystems). Cloned PCR products were sequenced in both directions using universal M13 primers and with internal primers in cases of products in excess of 800 bp. To minimize the number of PCR errors incorporated in cloned sequences, three plasmids from each cloning reaction were sequenced. In the consensus sequence of these replicates, any differences were changed to the common base found in other sequences of the same *Aly13* subtype from the same individual. In the alignment, we report the consensus sequence for a single individual per *Aly13* subtype (based on sequencing three clones or on direct sequencing using primers specific for a given subtype); we have observed minor variation within subtypes, which will be reported in a future publication.

Typing of *Aly13* subtypes in families: Identification of *Aly13* subtypes present in the progeny of the families used in the segregation analyses was based on the *S*-domain sequences, using a two-step procedure. First, the general primers described above were used to amplify any *Aly13* subtypes present in the individual. Restriction digestion of the resulting products then allowed us to generate an hypothesis about which *Aly13* subtypes were present in a particular individual through comparison with the restriction patterns for all known *Aly13* subtypes. This hypothesis was then tested using the primers specific for the hypothesized subtype. Putative subtypes identified in this manner were confirmed by sequencing. This method allowed all progeny from the families to be genotyped for the *Aly13* subtypes found in their parents.

Sequence analysis of *Aly13* subtypes: The *Aly13* subtypes

were aligned with one another and with Brassica *SRK* alleles using ClustalX (THOMPSON *et al.* 1997), followed by adjustments by eye using the SeAl 1.0 sequence editor (RAMBAUT 1998). The GenBank accession numbers for the Brassica *SRK* alleles are: AB013720, AB032473, AB032474, D30049, D38563, D38564, E15795, M76647, SEG_AB013717S, SEG_AB024419S, U00443, X79432, Y18259, Y18260, Z18921, and Z30211. To ensure that the reading frame was maintained when indels were incorporated, alignments were also checked by translating to amino acids. Manual adjustments were particularly important for the kinase domain because, as in Brassica, some of the introns have highly variable lengths and show only limited similarity among the *Aly13* subtypes. In particular, intron 1 (between the *S*-domain and the transmembrane domain) is extremely variable in length and very AT-rich and thus cannot be aligned with confidence. We therefore chose to present and discuss alignments of the *S*-domains and the kinase domains separately. Approximate intron-exon boundaries in the kinase domain were defined using *SRKa* (J. B. NASRALLAH and M. E. NASRALLAH, personal communication) aligned to a genomic clone of the same subtype (our *Aly13-13* subtype) and checked using GenBank descriptions of published Brassica *SRKs*. PAUP* 4.0b2 (SWOFFORD 2000) was used to calculate pairwise amino acid differences and nucleotide distances between *S*-domains of different *Aly13* subtypes. To visualize the sequence variability, nucleotide distances were calculated assuming the HKY85 substitution model, to take into account differences in base composition and transition/transversion ratios and to allow inclusion of sites with more than two variant bases (HASEGAWA *et al.* 1985). A tree of the *Aly13* and Brassica sequences was constructed, and the bootstrap support for branching relationships was evaluated using PAUP* and the HKY85 distance matrix. For construction of this tree we included as outgroups two *S*-domain genes that are unlinked to *SI*, *Ats1* from *A. thaliana* and *SLR1* from *B. oleracea* (GenBank accession nos.: *Ats1*, S84921; *SLR1*, Z26914). These genes have previously been suggested to be orthologues (SAKAMOTO *et al.* 1998). DnaSP 3.5 (ROJAS and ROJAS 1999) was used for construction of a sliding window of variability over the sequence alignments for *Aly13* and Brassica *SRK* alleles separately. The analyses were restricted to the seven *Aly13* subtypes where the longest sequence fragment was available (the subtypes included are indicated in Table 2; see below), and positions with indels were removed from the alignment (45 nucleotides in seven indels in total).

RESULTS

Crossing and segregation analysis

Evidence for linkage of *Aly13* subtypes with *SI* phenotypes: Figure 1, a–f, shows results from pollinations between full-sib progeny plants, which allowed us to score cross-incompatibility between siblings and to assign them to incompatibility groups within families. If a large majority of crosses involving several different progeny plants failed to produce fruits, the pollination was scored as incompatible. The *Aly13* sequence types (subtypes in the terminology defined above) of all individuals were then determined. The figures group progeny with the same *Aly13* sequence types, so that the correspondence between incompatibility reactions and *Aly13* subtypes within each sibship can be seen. Self-pollinations are not included; self-pollination rarely suc-

ceeds, except when a plant is severely stressed (M. H. SCHIERUP, personal observation).

Figure 1a shows results from pollinations between 15 full-sib plants from a cross between two plants originating from the North Carolina seed collection. The pollination results show that this sibship contains three incompatibility groups (note that groups I and II are phenotypically indistinguishable in their incompatibility reactions with plants of other groups). Within group IV, 27 of 98 pollinations produced fruits. These were mainly due to a single individual (10/25 pollinations produced fruits when it was used as the pollen donor, and 11/27 pollinations produced fruits when it was the recipient). This suggests that the strength of the SI response was reduced in this individual. Despite this, we can conclude that four incompatibility alleles must be segregating in this sibship. With respect to the *Aly13* sequences, the progeny fall into four groups, corresponding to those expected if the parental *Aly13* sequences (13-1/13-13 and 13-3/13-23) are allelic; no progeny plant inherits both (or neither) of the *Aly13* sequence types of either parent. Furthermore, the four genotypic classes based on the *Aly13* sequences correspond with the incompatibility groups.

This interpretation of the crossing results, with three incompatibility groups and four *Aly13* genotype groups, requires that some of the *S*-alleles are dominant to others, as is common in sporophytic SI systems (HATAKEYAMA *et al.* 1998). From the crossing table in Figure 1a, we can tentatively deduce some dominance relationships among the *S*-alleles in this family. We here assign these interpretations to the corresponding *Aly13* sequences, treating them as putative *S*-alleles. The observation that groups I and II are incompatible in both directions gives the dominance relations (where $>$ or $<$ mean, respectively, dominant and recessive, and $=$ means codominance): $13-23 \geq (13-1 \text{ and } 13-13)$. From the pollinations between group I + II and group III + IV, we can say that $13-3 < 13-1$ and/or $13-3 < 13-13$. Note that the results are the same in both reciprocals of any pollinations; *i.e.*, no sex-specific differences in dominance were found in this family.

The parents of the family shown in Figure 1b are also of North Carolina origin. Fourteen progeny plants were cross-pollinated in all combinations. There are four different incompatibility groups, corresponding exactly to the four genotypic classes expected from the *Aly13* sequence types found in the two parents. In the combinations scored as incompatible, a few fruits were produced, but these were mainly small (see footnotes in Figure 1, a–d and f). From this family, we further infer codominance of 13-23 with 13-20. The family in Figure 1c shares one parent (97F-13/5) with that in Figure 1a, while the other parent is from the Michigan population. Again, there are three incompatibility groups, again corresponding to the four genotypic classes. The dominance relationships can be reconciled with those deduced in

the family in Figure 1a as follows: $13-19 \geq (13-1 \text{ and } 13-13)$ and $(13-3 < 13-1)$ or $(13-3 < 13-13)$. A cross between two progeny from the family in Figure 1a (from groups II and III) produced the sibship shown in Figure 1d. Some pollinations were not done, but the data suggest two incompatibility groups, corresponding to the four genotypic classes defined by the *Aly13* sequences and suggesting the dominance relationships: $13-13 \geq (13-1 \text{ and } 13-3)$ and $13-23 \geq (13-1 \text{ and } 13-3)$. The family in Figure 1e is from a cross between an individual of Icelandic origin (98I-36/2) and an individual in group IV of the family in Figure 1a. There are only 10 progeny, and again the pollinations are incomplete. Two incompatibility groups were found, corresponding to four genotypic classes if we assume that 98I-36/2 carries a second *Aly13* subtype, 13-X, which has not yet been identified. The inferred dominance relationships in this sibship are as follows: $13-22 \geq (13-3 \text{ and } 13-13)$ and $13-X \geq (13-3 \text{ and } 13-13)$.

The family in Figure 1f is derived from a cross between two individuals from Michigan, which were not genotyped for *Aly13*. The 20 progeny fall into two incompatibility groups, with some partial compatibility within the second group. All progeny carry the *Aly13*-1 subtype, suggesting that one parent was homozygous for this subtype (proposed genotypes are given in the figure). Half the individuals carry *Aly13*-13, suggesting that the other parent was either 13-13/13-Y (where 13-Y is an unidentified *Aly13* subtype) or 13-1. With respect to dominance, this sibship suggests the relationships: $(13-13 > 13-1)$ and/or $(13-Y > 13-1)$.

Cross-pollination between two families: Figure 2 summarizes the results of cross-pollinations between progeny from the two families described in Figure 1, a and f. The incompatibility reaction between plants with the *Aly13* genotype 13-13/13-1 from one family and 13-13/13-3 from the other family supports our conclusion that the *Aly13*-13 subtype is associated with the same SI specificity in these two families of independent origin (North Carolina and Michigan, respectively). This sibship provides the following conclusions about dominance:

1. $13-13 < 13-23$
2. $(13-1 \leq 13-13)$ and $(13-3 \leq 13-13)$, with at least one of the inequalities being strict.

Segregation analysis: Given the high diversity of *Aly13* subtype sequences, it is important to test each sequence to see whether it could originate from a different locus. We therefore genotyped the parents and all progeny in five more families for putative alleles of *Aly13*. Table 1 summarizes the parental *Aly13* putative genotypes and the numbers of progeny with each possible *Aly13* sequence type. For all but one family, we were able to identify two *Aly13* subtypes in both parents. In all cases, each progeny plant inherited just one of the subtypes of each parent plant. In the remaining family, 00A4, *Aly13* subtypes in one parent were not identified, but

all progeny plants had just one of the two subtypes of the other parent. Thus, for these families, the evidence strongly suggests that the subtypes segregate as alleles at the same locus. Also, since some subtypes are present

in more than one of our families, the analyses can be combined for some subtypes. The families in Figure 1 show that subtypes 13-22, 13-13, and 13-3 segregate with the incompatibility groups within families, and Table 1

a

DONORS	RECIPIENTS				Aly13 sequences detected
	SI group (number of plants)				
	I (3)	II (4)	III (3)	IV (5)	
I	0/5	0/17	12/12	23/24	1,23
II	1/19	0/27	35/35	52/54	13,23
III	12/13	34/34	1/20	49/52	1,3
IV	19 ^a /22	45 ^a /50	55/55	27 ^b /98	3,13
Parents				97F 13-5	1,13
				97F 15-3	3,23

^a 1 small; ^b 9 small**b**

DONORS	RECIPIENTS				Aly13 sequences detected
	SI group (number of plants)				
	I (4)	II (5)	III (2)	IV (3)	
I	3 ^a /30	64/65	29/34	0/39	3,23
II	72/77	5 ^c /73	39/43	0/46	1,20
III	24/24	33/35	2 ^c /14	19/23	1,3
IV	10 ^b /74	5 ^d /51	21/23	3 ^d /18	20,23
Parents				97F 15-2	1,23
				97F 8-1	3,20

^a 2 small; ^b 6 small; ^c 5 small; ^d 3 small; ^e 1 small**c**

DONORS	RECIPIENTS				Aly13 sequences detected
	SI group (number of plants)				
	I (2)	II (1)	III (5)	IV (5)	
I	1 ^a /9	0/6	16/19	23 ^b /26	1,19
II	3 ^c /6	not done	6/6	7/8	13,19
III	16/17	6/6	3 ^a /65	32 ^a /37	1,3
IV	22 ^c /22	6/8	35/35	6 ^d /73	3,13
Parents				97F 13-5	1,13
				97F 12-3	3,19

^a 1 small; ^b 2 small; ^c 3 small; ^d 5 small**d**

DONORS	RECIPIENTS				Aly13 sequences detected
	SI group (number of plants)				
	I (8)	II (6)	III (3)	IV (2)	
I	0/13	0/18	16/17	not done	3,13
II	2 ^a /21	7 ^b /25	16/16	not done	1,13
III	8/8	8/9	0/6	not done	1,23
IV	3/3	not done	0/3	not done	3,23
Parents				98E 15-1	13,23
				98E 15-7	1,3

^a 2 small; ^b 4 small**e**

DONORS	RECIPIENTS				Aly13 sequences detected
	SI group (number of plants)				
	I (3)	II (3)	III (1)	IV (3)	
I	0/2	0/9	6/6	8/12	13,22
II	0/11	0/3	2/3	12/14	3,22
III	4/6	2/3	not done	0/6	3,X
IV	10/11	6/8	2/6	0/4	13,X
Parents				98I 36/2	22,X
				98E 15-11	3,13

f

DONORS	RECIPIENTS			Proposed genotypes
	SI group (number of plants)			
	I (10)	II (10)	Aly13 sequences detected	
I	6 ^a /72	66/70	1,13	A ₁₃ A ₁
II	60/68	22 ^b /59	1	A ₁ A _Y
Parents				A ₁ A ₁
				A _Y A ₁₃

^a 2 small; ^b 6 small

DONORS	RECIPIENTS					
	Genotype/Family					
	(Number of Individuals)					
	13-1/13-23 98E 15 (1)	13-13/13-23 98E 15 (1)	13-3/13-13 98E 15 (4)	13-1/13-3 98E 15 (2)	13-1/13-13 98E 17 (7)	13-1/13-Y 98E 17 (2)
13-1/13-23 98E 15	—	0/3	10/10	5/5	6/6	4/4
13-13/13-23 98E 15	0/3	—	15 ^a /16	6/6		
13-3/13-13 98E 15	13 ^a /16	15/18	22 ^c /72	30/30	5 ^c /15	6/6
13-1/13-3 98E 15	7/7	6/6	25/26	0/6	2/2	11/13
13-1/13-13 98E 17	3/4	3/3	5 ^c /13	8 ^a /10	4 ^c /46	52/53
13-1/13-Y 98E 17	9/9		4/4	12/14	41 ^a /44	13 ^c /36

^a 1 small; ^b 2 small; ^c 3 small; ^d 4 small; ^e 8 small

FIGURE 2.—Summary of pollinations performed between individuals of known genotype from the full-sib families described in Figure 1a (98E 15) and 1f (98E 17). The genotype, family of origin, and the number of individuals involved are indicated for recipients (x-axis) and donors (y-axis). Dashes indicate cross combinations that were not performed (because a genotype was represented by only a single individual). Numbers in each cell of the figure indicate the number of compatible pollinations out of the total tested, and shaded cells indicate incompatible mating groups. Footnotes give the numbers of crosses for which small fruits were produced (see text for explanation).

shows these to be allelic to 13-4, 13-5, 13-9, and 13-16.

Finally, the *S*-domain of one of our *Aly13* subtypes (*Aly13-13*) differs at just a single nonsynonymous site from that of the *SRKa* cDNA isolated by M. E. Nasrallah and J. B. Nasrallah and shown by them to be linked to SI in another independent family (KUSABA *et al.* 2001). Other subtypes that we have amplified have not yet been tested for linkage or segregation, but one sequence type (termed *Aly13-2* in CHARLESWORTH *et al.* 2000) has been found to be unlinked. This emphasizes the importance of linkage analysis before new sequences can be concluded to be alleles of our putative *Aly13* gene.

Summary of results on dominance relationships: Out of 11 different *Aly13* subtypes that behave as allelic, we obtained some evidence about the dominance relations for 7 that segregate with the incompatibility phenotype. If we assume that dominance is transitive (*i.e.*, that $A > B$ and $B > C$ implies that $A > C$), which conflicts with none of our crossing results, we can combine this evidence. We then have: (13-23 = 13-20, 13-19, 13-22) > (13-13) > (13-1, 13-3). There were no cases of differ-

ences in dominance in the stigma and pollen. Thus we have preliminary evidence for at least three dominance classes, but cannot resolve dominance within each class (except for codominance between 13-20 and 13-23). Furthermore, we cannot rule out the possibility of non-transitive relations, as observed in some cases in *Brassica* (THOMPSON and TAYLOR 1966).

Analysis of polymorphism: Table 2 summarizes the sequence data from the 11 subtypes from which linkage has been established from either pollination results or segregation analysis (see above and Table 2). As already mentioned, the *S*-domain of one of our *Aly13* subtypes (*Aly13-13*) differs at a single nonsynonymous site from that of a putative *A. lyrata SRK* allele (*SRKa*). This sequence was isolated from stigma cDNA. It has a kinase domain and is linked to SI (M. E. NASRALLAH and J. B. NASRALLAH, personal communication). A kinase domain was detected in 8 of the 11 subtypes analyzed here. For 1 of the remaining 3 subtypes (*Aly13-1*), a transcript without a kinase domain has been detected by rtPCR (M. H. SCHIERUP, unpublished results), suggesting that

FIGURE 1.—(a) Results from pollinations between plants from a cross between maternal parent plant 97F-13/5 and pollen donor plant 97F-15/3. To display the correspondence between incompatibility groups and *Aly13* genotypes, progeny are grouped into four sets (I–IV) within which all plants have the same *Aly13* sequences (shown in the right-hand column). The pollen donor groups are listed in the left-hand column, and those of the recipients in the pollinations are listed at the top of each column in the figure, with numbers of plants of each genotype in parentheses. In the grid are given the numbers of compatible pollinations/the total number of pollinations done between the donor and recipient groups of plants. If the two genotypes were scored as incompatible with one another, the square is shaded. The numbers of cases where only a small fruit developed are given in the footnotes. (b) Cross between individuals from North Carolina (98G 23 family). (c) Cross between a North Carolina (97F 13-5) female and a Michigan (97F 12-3) male. (d) Cross between individuals from the 98E 15 family. (e) Cross between an Icelandic individual (98I 36/2) and an individual from the 98E 15 family (North Carolina). (f) Cross between individuals from Michigan (98E 17 family). For details, see text.

TABLE 1
Segregation of parental *Aly13* subtypes in five families for which crossing data are not available

Family	00A1		00A2		00A3		00A4		00A5	
	Genotypes	No.	Genotypes	No.	Genotypes	No.	Genotypes	No.	Genotypes	No.
Female	99A-2/9	13-4/13-22	99A-2/3	13-5/13-16	99A-2/5	13-4/13-5	99A-4/7	99A-1/1	13-5/13-13	13-5/13-13
Male	99A-1/4	13-5/13-9	99A-1/3	13-5/13-13	99A-2/9	13-4/13-22	99A-1/1	99A-5/10	13-3/13-13	13-3/13-13
Progeny	13-22/13-5	6	13-5/13-16	4	13-4/13-22	2	13-5/?	13-5/13-13	5	
	13-4/13-9	2	13-16/13-13	2	13-4/13-4	6	13-13/?	13-5/13-3	5	
	13-4/13-5	2	13-5/13-13	5	13-5/13-4	3		13-13/13-3	4	
	13-9/13-2	2	13-5/13-5	2	13-5/13-22	2		13-13/13-13	9	

Family 00A1-00A3, crosses between individuals of Scottish origin; family 00A4, cross between North Carolina female (99A-4/7) and Scottish male (99A-1/1); family 00A5, cross between Scottish female (99A-1/1) and North Carolina male (99A-5/10).

Aly13-1 could be an *SLG* orthologue. However, because alternative transcripts are common in the Brassica system (TOBIAS and NASRALLAH 1996) this does not rule out the possibility that the *Aly13-1* genomic sequence includes a kinase domain. Due to the use of subtype-specific primers, our kinase domain sequences vary in length (last column of Table 2). Consequently, we focus on describing the polymorphism in the *S*-domain and show only preliminary findings on polymorphism within the kinase domain.

Polymorphism of the *Aly13 S*-domain sequences within *A. lyrata*: The alignment of the *Aly13* sequences to each other and to *SRKa* has six short indels in all, four of length three nucleotides and two of six nucleotides. There are no stop codons in any of the *Aly13* subtypes. In the alignment, the 12 conserved cysteine residues are at the same positions as in the Brassica *SRK* alleles (KUSABA *et al.* 1997). A block of amino acids WXSFXXTDT reported to be conserved in all plant receptor-like protein kinases (WALKER 1994) is also conserved in all *Aly13* subtypes. Because of the high diversity, we summarize relative divergence of alleles using simple measures. Table 3 shows pairwise differences in amino acid sequences (above diagonal, based on amino acid alignment) and nucleotide distances (below diagonal, using the HKY85 substitution model; HASEGAWA *et al.* 1985) for the 11 *Aly13* subtypes for which there is evidence of linkage. At the amino acid level, the mean pairwise difference was 42% (range 31–51% or 95–148 amino acid differences). At the nucleotide level, the average distance was 37% (range 23–47%). The subtypes are roughly equidistant from one another.

The two subtypes that appear to be associated with recessive SI phenotypes (*Aly13-1* and *Aly13-3*) are not particularly closely related (48% amino acid differences) and the distances from these subtypes to the other nine subtypes is on average 42%, very similar to the overall average. If the analysis is restricted to the eight linked *Aly13* subtypes known to have a kinase domain, the average amino acid difference is 40% (in the *S*-domain). Thus, there is no evidence that the three subtypes where a kinase domain has not been established cluster separately from the eight subtypes with kinase domains.

Comparison of the *Aly13 S*-domain sequences with Brassica *S*-alleles: The *S*-domains from 16 *SRK* alleles from the two closely related species *B. oleracea* and *B. rapa/campestris* were aligned with the 11 *Aly13* sequences discussed above. This introduces one more indel into the *Aly13* alignment because the Brassica *SRKs* are polymorphic for an indel of 12 bp that is not found among the *Aly13* sequences. Furthermore, one 3-bp insertion (AGG) at position 634 in the alignment is polymorphic in both *Aly13* and Brassica *SRK* data sets. For further analysis of the *S*-domain, the joint alignment was shortened to include only the 972 bp over which the *Aly13*

TABLE 2
Summary of data on 11 *Aly13* subtypes from *A. lyrata*

<i>Aly13</i> subtype	Evidence for linkage to SI phenotype	Kinase domain detected	Sequences used in sliding window plot	End of sequence
13-1	Yes	No	✓	S-domain
13-3	Yes	Yes	✓	Exon 5
13-4	Inferred from segregation	Yes	✓	Exon 7
13-5	Inferred from segregation	Yes	✓	Exon 4
13-9	Inferred from segregation	Yes	✓	Exon 5
13-13	Yes	Yes	✓	Exon 7
13-16	Inferred from segregation	Yes		Exon 6
13-19	Yes	Yes		Exon 6
13-20	Yes	Yes		Intron 4
13-22	Yes	No		S-domain
13-23	Yes	No	✓	S-domain

The evidence for the presence of a kinase domain (third column) is based on repeated attempted PCR amplifications (see MATERIALS AND METHODS). The fourth column indicates those sequences that were used to test whether the variability in the S-domain is in locations in the sequence similar to the locations of greatest variability in Brassica *SRK* alleles. The final column shows the approximate regions for which *A. lyrata* sequence data are available. The first six partial genomic sequences listed are all the same length and begin about 300 nucleotides from the beginning of the S-domain, but the other five S-domain sequences are shorter and vary in length.

subtypes were sequenced. An alignment of the *Aly13* sequences can be retrieved from the PopSet section of GenBank (with accession nos. AF328990–AF329000).

Figure 3 shows a gene tree reconstructed from the joint alignment of S-domain sequences and the *Ats1* and *SLRI* S-domain genes. Solid circles mark branches supported by >95% of bootstrap trials. The tree was not very well resolved (possibly due to recombination; see DISCUSSION) but illustrates the high diversity among alleles within and between the species. The Brassica class I and II *SRK* alleles form two distinct and well-supported groups, separated from all but one *Aly13* subtype (*Aly13-9*). There is no such clear division of the *Aly13* sequences into two clusters. This agrees with our crossing results (above) that showed no division of alleles into two distinct dominance classes. The *Ats1* and

SLRI sequences form a well-supported outgroup. If these genes are indeed orthologues (SAKAMOTO *et al.* 1998), then the tree suggests that this gene arose by duplication earlier than the diversification at the *SRK/Aly13* gene.

Table 4 compares the mean pairwise nucleotide distances and amino acid differences within the 11 *Aly13* sequences, within the 16 Brassica sequences, and between the species. For Brassica, the analysis was also done for 13 class I *SRK* alleles separately. The results show that diversity among the *Aly13* subtype sequences is higher than among the Brassica *SRKs*, even when Brassica class I and II alleles are combined. Furthermore, the average pairwise difference within *Aly13* subtypes is of similar magnitude to the average difference between sequences from *A. lyrata* and Brassica.

TABLE 3
S-domain diversity in 11 *Aly13* subtypes (pairwise differences)

Subtype	13-1	13-3	13-4	13-5	13-9	13-13	13-16	13-19	13-20	13-22	13-23
13-1		0.48	0.41	0.42	0.38	0.43	0.46	0.50	0.41	0.41	0.40
13-3	0.44		0.45	0.39	0.44	0.41	0.46	0.44	0.44	0.51	0.44
13-4	0.36	0.44		0.40	0.32	0.44	0.44	0.42	0.38	0.38	0.35
13-5	0.39	0.33	0.39		0.39	0.31	0.35	0.43	0.40	0.48	0.40
13-9	0.32	0.40	0.27	0.36		0.42	0.44	0.46	0.39	0.38	0.33
13-13	0.37	0.36	0.41	0.23	0.38		0.38	0.44	0.42	0.48	0.42
13-16	0.44	0.47	0.46	0.27	0.41	0.27		0.41	0.43	0.44	0.45
13-19	0.44	0.40	0.38	0.37	0.38	0.38	0.36		0.40	0.44	0.44
13-20	0.34	0.41	0.33	0.35	0.35	0.38	0.39	0.37		0.41	0.38
13-22	0.34	0.42	0.33	0.40	0.32	0.40	0.40	0.35	0.35		0.40
13-23	0.33	0.40	0.33	0.38	0.32	0.40	0.45	0.39	0.33	0.32	

Diversity was calculated using PAUP 4.0b2 (SWOFFORD 2000). Above diagonal, uncorrected amino acid differences; below diagonal, nucleotide distances (HKY85).

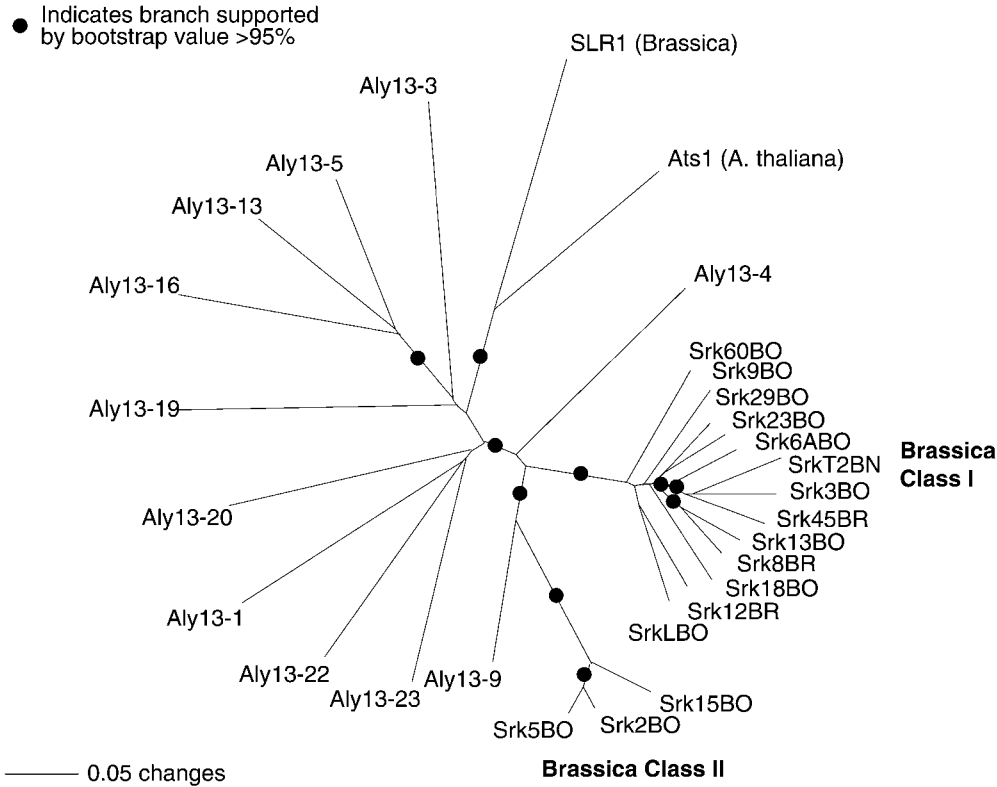


FIGURE 3.—Unrooted gene tree based on the S-domains of 11 *Aly13* subtypes, 16 Brassica *SRK* alleles, *Ats1*, and *SLR1* (constructed using PAUP* 4.0). Neighbor joining was used, with a distance matrix calculated from nucleotide sequences using the HKY85 substitution model. Solid circles on branches indicate that the branch is supported by >95% of bootstrap replicates.

Figure 4 shows nucleotide diversity in a sliding window over the seven *Aly13* subtypes from which the most complete sequences were obtained (solid line) and the 16 Brassica *SRK* alleles (dashed line). The pattern of diversity is similar in the two species, with peaks of diversity at similar positions as the three hypervariable regions in Brassica (SIMS 1993; NISHIO and KUSABA 2000). To test whether the pattern of variability is significantly correlated in the sets of sequences from the two species, we used the first six sequences listed in Table 2. All sites at which there were no variants (*i.e.*, where both species are fixed for the same residue) were removed to avoid biasing the result by including nonvariable sites. Such sites are evidently correlated in diversity in the two species and may be at the same sites in both, but provide no information about differences between regions of moderate and extremely high variability. We then tested

the diversity values of the remaining sites by Spearman rank correlation. For the six *Aly13* sequences that include the sites that are hypervariable in Brassica, 278 sites were removed, and the correlation was significant ($P = 0.02$) for the remaining 622 sites; a similar analysis, based only on amino acid replacement sites, was also highly significant ($P < 0.001$).

Table 5 shows evidence that the kinase domain is also highly polymorphic and compares diversity in five *Aly13* subtypes (13-3, 13-9, 13-13, 13-16, and 13-19) with the corresponding values for Brassica *SRK* sequences. This was done for each available intron and exon separately using the *SRKa* cDNA to locate putative splicing positions and confirmed using data from GenBank for Brassica *SRK* alleles. The sequences were of unequal lengths (Table 2), but at least five *Aly13* subtypes were analyzed for each exon or intron. Because most Brassica *SRK*

TABLE 4

Average nucleotide distances (HKY85 substitution model) and amino acid differences between the sequences of the S-domains of *A. lyrata* *Aly13* subtypes and Brassica *SRK* alleles

	<i>A. lyrata</i> alleles	Brassica Class I + II	Brassica Class I	Between Brassica and <i>A. lyrata</i>
No. of sequences	11	16	13	27
Mean nucleotide distances	0.37	0.21	0.15	0.38
Mean amino acid differences	0.42	0.26	0.22	0.41

The last column shows the average differences between Brassica and *A. lyrata* sequences using our joint alignment.

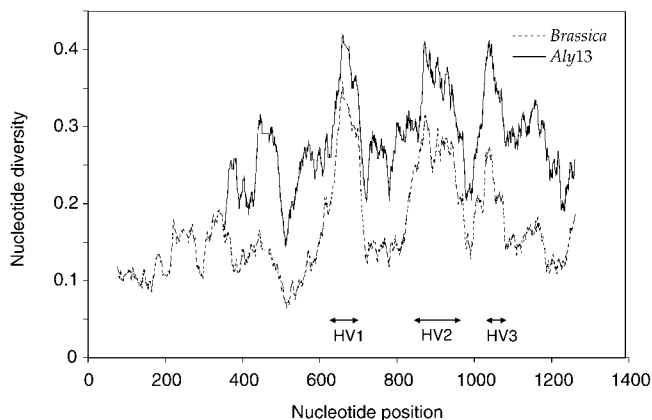


FIGURE 4.—Sliding window analysis of nucleotide diversity (π) for seven *Aly13* subtypes (solid line) and 16 *Brassica SRK* alleles (dashed line). The positions refer to the nucleotide position in the alignment of the *Brassica SRK* alleles. The window size is 50 nucleotides with a step size of one nucleotide. The arrows indicate the positions of the three hypervariable regions in *Brassica*.

sequences are from cDNA, the *Brassica* intron values in Table 5 are based on four Class I sequences only. Table 5 shows that *Aly13* subtypes have higher diversity than *Brassica SRKs* in the kinase domain, as well as in the *S*-domain (see above).

DISCUSSION

Our results suggest that the SI system of *A. lyrata* shares a common evolutionary origin with that in *Bras-*

sica. This is not surprising. It is observed generally that the incompatibility genes in different members of the same family are homologous (see UYENOYAMA 1995). However, the allelic variants are highly diverged in sequence, implying that the age of the alleles is very high, which is again not unexpected since the cDNA sequences of different gametophytic *S*-alleles also differ greatly (e.g., RICHMAN *et al.* 1996). The overall diversity in the *S*-domains of *Aly13* subtypes is, however, larger than the diversity of *Brassica SRK* alleles, even when alleles of class I and II are combined. The diversity among *Aly13* subtype sequences is close to that observed in the gametophytic self-incompatibility system of species of Solanaceae (RICHMAN *et al.* 1996), which has previously been considered more variable than the sporophytic system of Brassicaceae (SIMS 1993; UYENOYAMA 1995).

We cannot rule out the possibility that some of our subtypes are sequences of *S*-domain genes that are linked to the self-incompatibility genes, but not involved in determining the SI specificity. However, this is unlikely for most of the putative alleles, for the following reasons (in addition to the segregation data given above). First, primers specific for given subtypes never detect a subtype in a large proportion of individuals, as should happen if a sequence represents a separate locus (unless this locus is highly polymorphic). Second, except for *Aly13-2*, which is unlinked to the *S*-locus, we didn't find more than two subtypes in any given individual. It is therefore unlikely that other closely linked genes are

TABLE 5

Comparison of diversity within *Aly13* subtypes in *A. lyrata* and *SRK* alleles of *Brassica*

	<i>A. lyrata</i> Subtypes 3, 9, 13, 16, and 19		<i>Brassica</i> Class I and II	
	Length range of regions	Diversity	Length range of regions	Diversity
S-domain	963	0.29	963	0.18
Intron 1	>500	Unalignable	>500	Unalignable
Transmembrane	108–138	0.29	129–147	0.17
Intron 2	77–147	Unalignable	105–192	0.47
Exon 3	176–197	0.32	182–200	0.16
Intron 3	65–117	0.48	74–80	0.26
Exon 4	211	0.20	208–211	0.15
Intron 4	84–102	0.42	89	0.14
Exon 5	238–241 ^a	0.16	238	0.12
Intron 5			82–115	0.37
Exon 6			130–151	0.16
Intron 6			79–108	0.36
Exon 7			318–342	0.16

Average nucleotide diversities (uncorrected P-distances; SWOFFORD 2000) are shown. Length ranges are the minimum and maximum length of the given exon/intron for the sequences used in each species. Values for *A. lyrata* are based on five *Aly13* subtypes. Intron 1, intron 2, and the first 9 bp of exon 3 are unalignable among the *Aly13* subtypes and were excluded from the analysis. Exon values for *Brassica* are based on 16 *SRK* alleles from *B. rapa* and *B. oleracea* (13 Class I and 3 Class II alleles), and intron values are based on 4 *SRK* alleles (all Class I).

^a The last 58 bp were excluded in calculations because 13-3 and 13-9 sequences ended there.

often amplified and contribute to our sequences. If so, they must be present in only a small minority of individuals, or else be highly polymorphic, so that the primers often fail to amplify. Further evidence that all the *Al*13 subtypes reported here represent allelic variants at a single *SRK* locus will require cloning of a large genomic region flanking the gene for each subtype, and expression studies will be needed to test expression in stigmas.

Among the 11 *Al*13 subtypes for which we have evidence of linkage to the *S*-locus, our partial information on dominance relationships suggests at least three dominance levels with some codominance evident between members of the same level. We did not observe any sex differences in dominance among these 7 subtypes, but other crossing results have shown that such interactions do occur in *A. lyrata* (SCHIERUP 1998; MABLE *et al.*, unpublished results).

The phylogeny of *S*-domain sequences suggests that the polymorphism within Brassica classes I and II *SRK* alleles arose after the split of the two species. However, the opposite conclusion is suggested by the observation that an indel polymorphism is present in both species (although these nucleotides may have been inserted or deleted more than once). In addition, the *Al*13 subtype 13-9 appears to be more closely related to the Brassica class II alleles than to any of the other sequence types. We do not yet have dominance information for this allele because its linkage to the *S*-locus was inferred from segregation analysis with respect to other putative alleles that are linked, and pollination data have not been obtained. It will be interesting to test whether this allele is recessive in pollen, like the Brassica class II alleles. However, the conclusions from phylogenetic analyses must be treated cautiously when using a locus such as *SRK* that is a member of a multi-gene family in which recombination or gene conversion between alleles at different loci may occur. If such exchanges occur within either species, the gene tree will not represent the ancestry of the alleles, but will have a complex network structure (HUDSON 1990). There is evidence suggesting the occurrence of recombination in the Brassica *SLG* alleles (AWADALLA and CHARLESWORTH 1999). With sequence data from alleles from natural populations, it will be possible to test this in *A. lyrata* also. At present, we note that the *Al*13 subtype sequences are roughly equidistant from one another (Figure 3), which is expected if recombination is common (SCHIERUP and HEIN 2000).

Our analysis of the few *Al*13 kinase domains so far sequenced shows that these domains are highly variable. This is surprising if the *S*-locus region undergoes recombination. If the amino acid residues that are involved in recognition, and subject to balancing selection, are in the hypervariable regions of the *S*-domain, one would expect no further peaks of diversity as far away as the kinase domain (ANDOLFATTO and NORDBORG 1998). Our preliminary analysis suggests that diversity at both

silent sites in the exons, and at sites in the intron sequences, decreases with increasing distance from the *S*-domain (Table 5). This resembles the pattern in certain major histocompatibility complex genes, where gene conversion has been suggested as an explanation for decreased diversity at sites distant from the peptide-binding region, which is probably under balancing selection (BERGSTROM *et al.* 1998). However, the distances between sites in those genes are shorter than in *SRK* genes, which have a large intron after the end of the *S*-domain (NASRALLAH *et al.* 1988 and our own unpublished data from *A. lyrata*).

Finally, our finding that most of the *Al*13 subtypes contain a kinase domain suggests that *A. lyrata* may not have a haplotype structure of linked *SRK* and *SLG* alleles with similar *S*-domains. If there were haplotype structure similar to that of Brassica, we should find pairs of closely related sequences segregating with the different specificities, and three or four *Al*13 subtypes might be observed in at least some individuals. This does not seem to be found, which raises several possibilities. First, it is possible that *SLG* is indeed not necessary for the determination of the SI phenotypes in *A. lyrata* and that it is absent from most or all haplotypes. Second, and alternatively, pairs of *SRK/SLG* genes may have (almost) identical *S*-domains, so that we are unable to distinguish the two loci by the methods used here. Third, it is possible that *SRK* and *SLG* orthologues in *A. lyrata* are so diverged that we are able to detect only the *SRK* and not the *SLG* orthologue by our PCR-based approach. Work is in progress to test between these possibilities.

We are very grateful to M. E. Nasrallah and J. B. Nasrallah for sharing information about the sequences of putative *S*-alleles of *A. lyrata*. The first attempts at isolating *S*-domain genes were done in Chuck Langley's Lab, UC Davis, and we thank Chuck for his support and for discussions about SI. We thank the staff at the University of Edinburgh and Paaskehojgard Experimental Station for growing the plants, and the following people for seeds used in this work: T. E. Thorhallsdottir, C. H. Langley, and R. Mauricio. This work was supported by the Biotechnology and Biological Sciences Research Council of the United Kingdom, including support for B. K. Mable. M. H. Schierup was supported by the Danish Natural Sciences Research Council (grant nos. 9701412 and 1262), and wishes to thank Camilla Haakonsen for excellent lab work. D. Charlesworth was supported by the Natural Environment Research Council of Great Britain and Edinburgh University, and P. Awadalla by an Edinburgh University Faculty of Science and Engineering Scholarship.

LITERATURE CITED

- ANDOLFATTO, P., and M. NORDBORG, 1998 The effect of gene conversion on intralocus associations. *Genetics* **148**: 1397-1399.
- AWADALLA, P., and D. CHARLESWORTH, 1999 Recombination and selection at *Brassica* self-incompatibility loci. *Genetics* **152**: 413-425.
- BERGSTROM, T. F., A. JOSEFSSON, H. A. ERLICH and U. GYLLENSTEN, 1998 Recent origin of *HLA-DRB1* alleles and implications for human evolution. *Nat. Genet.* **18**: 237-242.
- BI, Y. M., N. BRUGIERE, Y. CUI, D. R. GORING and S. J. ROTHSTEIN, 2000 Transformation of *Arabidopsis* with a *Brassica SLG/SRK*

- region and *ARCI* gene is not sufficient to transfer the self-incompatibility phenotype. *Mol. Gen. Genet.* **263**: 648–654.
- CHARLESWORTH, D., 2000 How can two-gene models of self-incompatibility generate new specificities? *Plant Cell* **12**: 309–310.
- CHARLESWORTH, D., P. AWADALLA, B. K. MABLE and M. H. SCHIERUP, 2000 Population-level studies of multiallelic self-incompatibility loci, with particular reference to Brassicaceae. *Ann. Bot.* **85**: 227–239.
- CONNER, J. A., P. CONNER, M. E. NASRALLAH and J. B. NASRALLAH, 1998 Comparative mapping of the *Brassica* *S*-locus region and its homolog in *Arabidopsis*: implications for the evolution of mating systems in the Brassicaceae. *Plant Cell* **10**: 801–812.
- CUI, Y. H., Y. M. BI, N. BRUGIERE, M. A. ARNOLDO and S. J. ROTHSTEIN, 2000 The *S* locus glycoprotein and the *S* receptor kinase are sufficient for self-pollen rejection in *Brassica*. *Proc. Natl. Acad. Sci. USA* **97**: 3713–3717.
- DWYER, K. G., M. A. BALENT, J. B. NASRALLAH and M. E. NASRALLAH, 1991 DNA sequences of self-incompatibility genes from *Brassica campestris* and *B. oleracea*: —polymorphism predating speciation. *Plant Mol. Biol.* **16**: 481–486.
- HASEGAWA, M., H. KISHINO and T. YANO, 1985 Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. *J. Mol. Evol.* **22**: 160–174.
- HATAKEYAMA, K., M. WATANABE, T. TAKASAKI, K. OJIMA and K. HINATA, 1998 Dominance relationships between *S*-alleles in self-incompatible *Brassica campestris* L. *Heredity* **80**: 241–247.
- HUDSON, R. R., 1990 Gene genealogies and the coalescent process. *Oxf. Surv. Evol. Biol.* **7**: 1–45.
- HUGHES, A. L., and M. YEAGER, 1998 Natural selection at major histocompatibility complex loci of vertebrates. *Annu. Rev. Genet.* **32**: 415–435.
- JUNGHANS, H., and M. METZLAFF, 1990 A simple and rapid method for the preparation of total plant DNA. *Biotechniques* **8**: 176.
- KARPECHENKO, G. D., 1927 Polyplloid hybrids of *Raphanus sativus* X *Brassica oleracea*. *Appl. Bot. Genet. Plant Breeding* **17**: 305–410.
- KUSABA, M., T. NISHIO, Y. SATTA, K. HINATA and D. OCKENDON, 1997 Striking sequence similarity in inter- and intra-specific comparisons of class I *SLG* alleles from *Brassica oleracea* and *Brassica campestris*: implications for the evolution and recognition mechanism. *Proc. Natl. Acad. Sci. USA* **94**: 7673–7678.
- KUSABA, M., K. DWYER, J. HENDERSHOT, J. VREBALOV, J. B. NASRALLAH *et al.*, 2001 Self-incompatibility in the genus *Arabidopsis*. Characterization of the *S* locus in the outcrossing *A. lyrata* and its autogamous relative *A. thaliana*. *Plant Cell* **13**: 627–643.
- MAHY, G., and A. L. JACQUEMART, 1999 Early inbreeding depression and pollen competition in *Calluna vulgaris* (L.) Hull. *Ann. Bot.* **83**: 697–704.
- MCNAUGHTON, I. H., 1973 Resistance of *Raphanobrassica* to clubroot disease. *Nature* **243**: 547–548.
- NASRALLAH, J. B., T.-H. KAO, T.-H. CHEN, M. GOLDBERG and M. E. NASRALLAH, 1987 Amino-acid sequence of glycoproteins encoded by three alleles of the *S*-locus of *Brassica oleracea*. *Nature* **326**: 617–619.
- NASRALLAH, J. B., S.-M. YU and M. E. NASRALLAH, 1988 Self-incompatibility genes of *Brassica oleracea*: expression, isolation and structure. *Proc. Natl. Acad. Sci. USA* **85**: 5551–5555.
- NASRALLAH, J. B., T. NISHIO and M. E. NASRALLAH, 1991 The self-incompatibility genes of *Brassica*: expression and use in genetic ablation of floral tissues. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **42**: 393–422.
- NISHIO, T., and M. KUSABA, 2000 Sequence diversity of *SLG* and *SRK* in *Brassica oleracea* L. *Ann. Bot.* **85** (Suppl. A): 141–146.
- NORDBORG, M., B. CHARLESWORTH and D. CHARLESWORTH, 1996 The effect of recombination on background selection. *Genet. Res.* **67**: 159–174.
- RAMBAUT, A., 1998 Se-AL: Sequence Alignment Editor v1.0 alpha 1. Distributed over the World Wide Web (<http://evolve.zps.ox.ac.uk>).
- RICHMAN, A. D., M. K. UYENOYAMA and J. R. KOHN, 1996 Allelic diversity and gene genealogy at the self-incompatibility locus in the Solanaceae. *Science* **273**: 1212–1216.
- ROJAS, J., and R. ROJAS, 1999 DnaSP version 3: an integrated program for molecular population genetics and molecular evolution analysis. *Bioinformatics* **15**: 174–175.
- ROLLINS, R. C., 1993 *The Cruciferae of Continental North America*. Stanford University Press, Stanford, CA.
- SAKAMOTO, K., M. KUSABA and T. NISHIO, 1998 Polymorphism of the *S*-locus glycoprotein gene (*SLG*) and the *S*-locus related gene (*SLR1*) in *Raphanus sativus* L. and self-incompatible ornamental plants in the Brassicaceae. *Mol. Gen. Genet.* **258**: 397–403.
- SCHIERUP, M. H., 1998 The effect of enzyme heterozygosity on growth in a strictly outcrossing species, the self-incompatible *Arabidopsis petraea* (Brassicaceae). *Heredity* **128**: 21–31.
- SCHIERUP, M. H., and J. HEIN, 2000 Consequences of recombination on traditional phylogenetic analysis. *Genetics* **156**: 879–891.
- SCHIERUP, M. H., X. VEKEMANS and F. B. CHRISTIANSEN, 1998 Allelic genealogies in sporophytic self-incompatibility systems in plants. *Genetics* **150**: 1187–1198.
- SCHIERUP, M. H., X. VEKEMANS and D. CHARLESWORTH, 2000 The effect of subdivision on variation at multi-allelic loci under balancing selection. *Genet. Res.* **76**: 51–62.
- SCHOPFER, C. R., M. E. NASRALLAH and J. B. NASRALLAH, 1999 The male determinant of self-incompatibility in *Brassica*. *Science* **286**: 1697–1700.
- SHIBA, H., N. KIMURA, S. TAKAYAMA, K. HINATA, A. SUZUKI *et al.*, 2000 Alteration of the self-incompatibility phenotype in *Brassica* by transformation of the antisense *SLG* gene. *Biosci. Biotechnol. Biochem.* **64**: 1016–1024.
- SIMS, T. L., 1993 Genetic regulation of self-incompatibility. *Crit. Rev. Plant Sci.* **12**: 129–167.
- STEIN, J., B. HOWLETT, D. C. BOYES, M. E. NASRALLAH and J. B. NASRALLAH, 1991 Molecular cloning of a putative receptor protein kinase gene encoded at the self-incompatibility locus of *Brassica oleracea*. *Proc. Natl. Acad. Sci. USA* **88**: 8816–8820.
- STROBECK, C., 1983 Expected linkage disequilibrium for a neutral locus linked to a chromosomal arrangement. *Genetics* **103**: 545–555.
- SWOFFORD, D. L., 2000 *PAUP*: Phylogenetic Analysis Using Parsimony (*and Other Methods)*, Version 4. Sinauer Associates, Sunderland, MA.
- TAKAHATA, N., 1990 A simple genealogical structure of strongly balanced allelic lines and trans-species evolution of polymorphism. *Proc. Natl. Acad. Sci. USA* **87**: 2419–2423.
- TAKAHATA, N., and Y. SATTA, 1998 Footprints of intragenic recombination at HLA loci. *Immunogenetics* **47**: 430–441.
- TAKASAKI, T., K. HATAKEYAMA, G. SUZUKI, M. WATANABE, A. ISOGAI *et al.*, 2000 The Receptor kinase determines self-incompatibility in *Brassica stigma*. *Nature* **403**: 913–916.
- TAKAYAMA, S., H. SHIBA, M. IWANO, H. SHIMOSATO, F. S. CHE *et al.*, 2000 The pollen determinant of self-incompatibility in *Brassica campestris*. *Proc. Natl. Acad. Sci. USA* **97**: 1920–1925.
- THOMPSON, J. D., T. J. GIBSON, F. PLEWNIAC, F. JEANMOUGIN and D. G. HIGGINS, 1997 The ClustalX windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res.* **24**: 4876–4882.
- THOMPSON, K. F., and J. P. TAYLOR, 1966 Non-linear dominance relationships between *S* alleles. *Heredity* **21**: 345–362.
- TOBIAS, C. M., and J. B. NASRALLAH, 1996 An *S*-locus-related gene in *Arabidopsis* encodes a functional kinase and produces two classes of transcripts. *Plant J.* **10**: 523–531.
- UYENOYAMA, M. K., 1995 A generalized least-squares estimate for the origin of sporophytic self-incompatibility. *Genetics* **139**: 975–992.
- VEKEMANS, X., and M. SLATKIN, 1994 Gene and allelic genealogies at a gametophytic self-incompatibility locus. *Genetics* **137**: 1157–1165.
- WALKER, J. C., 1994 Structure and function of the receptor-like protein-kinases of higher-plants. *Plant Mol. Biol.* **26**: 1599–1609.
- WATANABE, M., A. ITO, Y. TAKADA, C. NINOMIYA, T. KAKIZAKI *et al.*, 2000 Highly divergent sequences of the pollen self-incompatibility (*S*) gene in class-I *S* haplotypes of *Brassica campestris* (syn. *rapa*) L. *FEBS Lett.* **473**: 139–144.
- YANG, Y. W., K. N. LAI, P. Y. TAI, D. P. MA and W. H. LI, 1999a Rates of nucleotide substitution in angiosperm mitochondrial DNA sequences and dates of divergence between *Brassica* and other angiosperm lineages. *J. Mol. Evol.* **48**: 597–604.
- YANG, Y. W., K. N. LAI, P. Y. TAI, D. P. MA and W. H. LI, 1999b Molecular phylogenetic studies of *Brassica*, *Rorippa*, *Arabidopsis* and allied genera based on the internal transcribed spacer region of 18S–25S rDNA. *Mol. Phylogenet. Evol.* **13**: 455–462.
- YU, K., T. L. GLAVIN, D. F. GÖRING and S. J. ROTHSTEIN, 1996 Molecular characterization of the *S* locus in two self-incompatible *Brassica napus* lines. *Plant Cell* **8**: 2369–2380.